

#2  
Docket No. 1081.1055/JDE

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

71058 U.S. PTO  
08/932427  
09/17/97

In re Patent Application of:

Yoshio KITAMURA

Group Art Unit: To Be Assigned

Serial No.: To Be Assigned

Examiner: To Be Assigned

Filed: September 17, 1997

For: RAID APPARATUS AND ACCESS CONTROL METHOD THEREFOR

**SUBMISSION OF CERTIFIED COPY OF PRIOR FOREIGN  
APPLICATION IN ACCORDANCE  
WITH THE REQUIREMENTS OF 37 C.F.R. §1.55**

*Assistant Commissioner for Patents  
Washington, D.C. 20231*

*Sir:*

In accordance with the provisions of 37 C.F.R. §1.55, Applicants submit herewith a certified copy of the following foreign application:

Japanese Patent Application No. 9-004427, filed January 14, 1997.

It is respectfully requested that Applicants be given the benefit of the foreign filing date as evidenced by the certified papers attached hereto, in accordance with the requirements of 35 U.S.C. §119.

Respectfully submitted,

STAAS & HALSEY

Dated: September 17, 1997

By:

  
\_\_\_\_\_  
James D. Halsey, Jr.  
Registration No. 22,729

700 Eleventh Street, N.W.  
Suite 500  
Washington, D.C. 20001  
(202) 434-1500

日 本 国 特 許 庁

PATENT OFFICE  
JAPANESE GOVERNMENT

71058 U.S. PTO  
08/932427  
09/17/97

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

1 9 9 7 年 1 月 1 4 日

出 願 番 号

Application Number:

平成 9 年特許願第 0 0 4 4 2 7 号

出 願 人

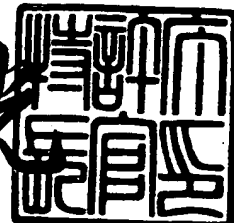
Applicant (s):

富士通株式会社

1 9 9 7 年 6 月 6 日

特 許 庁 長 官  
Commissioner,  
Patent Office

荒井 寿光



【書類名】 特許願

【整理番号】 9607010

【提出日】 平成 9年 1月14日

【あて先】 特許庁長官 荒井 寿光 殿

【国際特許分類】 G06F 13/12

【発明の名称】 R A I D装置及びそのアクセス制御方法

【請求項の数】 10

【発明者】

    【住所又は居所】 神奈川県横浜市港北区新横浜2丁目4番19号 株式会社富士通プログラム技研内

    【氏名】 北村 嘉朗

【特許出願人】

    【識別番号】 000005223

    【氏名又は名称】 富士通株式会社

    【代表者】 関澤 義

【代理人】

    【識別番号】 100094514

    【弁理士】

    【氏名又は名称】 林 恒▲徳▼

【代理人】

    【識別番号】 100094525

    【弁理士】

    【氏名又は名称】 土井 健二

【手数料の表示】

    【予納台帳番号】 041380

    【納付金額】 21,000円

【提出物件の目録】

    【物件名】 明細書 1

    【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9001093

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 R A I D装置及びそのアクセス制御方法

【特許請求の範囲】

【請求項1】 一の論理ボリュームを構成するための複数の物理ディスクユニットと、指定された論理ボリュームを構成する物理ディスクユニットをアクセスして、前記指定された論理ボリュームをアクセスするディスクコントローラとを有し、

前記ディスクコントローラは、

各物理ディスクユニット毎に、各物理ディスクユニットに要求されているオペレーションの数を格納するメモリと、

前記指定された論理ボリュームを構成する前記複数の物理ディスクユニットを選択し、且つ前記選択された複数の物理ディスクユニットの内、前記オペレーションの数に応じて物理ディスクユニットをアクセスするための制御手段とを有することを

特徴とするR A I D装置。

【請求項2】 請求項1のR A I D装置において、

前記制御手段は、

上位とのインターフェース制御を行うためのチャネルアダプター回路と、

依頼されたオペレーションに応じて、前記物理ディスクユニットをアクセスするためのデバイスアダプター回路と、

前記チャネルアダプター回路からの転送要求に応じて、前記メモリを参照して、前記オペレーションの数に応じて物理ディスクユニットを決定し、且つ前記決定した物理ディスクユニットへのアクセスのためのオペレーションを前記デバイスアダプター回路に依頼するリソースマネージャー回路とを有することを

特徴とするR A I D装置。

【請求項3】 請求項2のR A I D装置において、

前記リソースマネージャー回路は、

前記オペレーションの依頼に応じて、前記メモリの決定された物理ディスクユニットに対応するオペレーションの数を加算し、且つ前記デバイスアダプター回

路のオペレーションの終了に応じて、前記メモリのオペレーションの終了した物理ディスクユニットに対応するオペレーションの数を減算することを

特徴とするRAID装置。

【請求項4】 請求項2又は3のRAID装置において、

前記メモリは、前記物理ディスクユニットの状態を示すステータス情報を格納し、

前記リソースマネージャ回路は、前記ステータス情報を参照して、前記論理ボリュームを構成する各々の配置された物理ディスクユニットの正常を判定し、正常な物理ディスクユニットをアクセスの対象として、選択することを

特徴とするRAID装置。

【請求項5】 請求項2又は3又は4のRAID装置において、

前記メモリは、前記論理ボリューム毎に、前記論理ボリュームを格納する複数の物理ディスクユニットの情報を格納し、

前記リソースマネージャ回路は、前記メモリを参照して、前記論理ボリュームが配置された物理ディスクユニットを選択することを

特徴とするRAID装置。

【請求項6】 一の論理ボリュームを構成するための複数の物理ディスクユニットと、指定された論理ボリュームを構成する物理ディスクユニットをアクセスして、前記指定された論理ボリュームをアクセスするディスクコントローラとを有するRAID装置のアクセス制御方法において、

前記指定された論理ボリュームを構成する前記複数の物理ディスクユニットを選択するステップと、

前記選択された複数の物理ディスクユニットの内、前記物理ディスクユニットに要求されているオペレーションの数に応じて物理ディスクユニットをアクセスするステップとを有することを

特徴とするRAID装置のアクセス制御方法。

【請求項7】 請求項6のRAID装置のアクセス制御方法において、

チャネルアダプター回路により上位から転送要求を受信するステップと、

リソースマネージャ回路により、前記転送要求に応じて、前記オペレーショ

ンの数に応じて物理ディスクユニットを決定し、且つ前記決定した物理ディスクユニットへのアクセスのためのオペレーションを発するステップと、

デバイスアダプター回路により、前記依頼されたオペレーションに応じて、前記物理ディスクユニットをアクセスするステップとを有することを

特徴とするRAID装置のアクセス制御方法。

【請求項8】 請求項7のRAID装置のアクセス制御方法において、

前記オペレーションの依頼に応じて、メモリに格納された前記決定された物理ディスクユニットに対応するオペレーションの数を加算するステップと、

前記物理ディスクユニットのアクセスのオペレーションの終了に応じて、前記メモリのオペレーションの終了した物理ディスクユニットに対応するオペレーションの数を減算するステップとを更に有することを

特徴とするRAID装置のアクセス制御方法。

【請求項9】 請求項7又は8のRAID装置のアクセス制御方法において

前記リソースマネージャ回路により、メモリに格納された前記物理ディスクの状態を示すステータス情報を参照して、前記論理ボリュームが配置された物理ディスクユニットの正常を判定するステップと、

前記正常な物理ディスクユニットをアクセスの対象として、選択するステップとを有することを

特徴とするRAID装置のアクセス制御方法。

【請求項10】 請求項7又は8又は9のRAID装置のアクセス制御方法において、

前記選択するステップは、前記リソースマネージャ回路により、メモリの論理ボリュームが配置された複数の物理ディスクユニットの情報を参照して、前記論理ボリュームが配置された物理ディスクユニットを選択するステップであることを

特徴とするRAID装置のアクセス制御方法。

【発明の詳細な説明】

【0001】

## 【発明の属する技術分野】

本発明は、同一の論理ボリュームを複数の物理ディスクユニットに配置した R A I D 装置及びそのアクセス制御方法に関し、特に、アクセス要求の偏りを防止するための R A I D 装置及びそのアクセス制御方法に関する。

## 【0002】

コンピュータシステムにおける外部記憶システムとして、磁気ディスクシステム等のディスク記憶システムが採用されている。コンピュータは、ディスク記憶システムに対し、O S (Operating System) が認識する論理ボリューム名で、ディスク記憶システムをアクセスする。一方、ディスク記憶システムには、論理ボリュームが配置される。

## 【0003】

このようなディスク記憶システムにおいて、各論理ボリュームを構成する物理ディスクユニットを1つにすることは、その論理ボリュームを構成する物理ディスクユニットが故障した時に、その論理ボリュームは使用できなくなる。

## 【0004】

これを防止するために、R A I D (Redundant Arrays of Inexpensive Disks) 装置が提案されている。R A I D 装置では、一の論理ボリュームが、異なる複数の物理ディスクユニットに配置される。そして、一のディスクユニットが故障した時には、当該ディスクユニットに格納されていた情報と同一の情報を有する同一の論理ボリュームに配置された他のディスクユニットを使用する。これにより、ディスクユニットの故障により、論理ボリュームが使用できなくなる事態を防止できる。

## 【0005】

## 【従来の技術】

図8は、従来技術の説明図である。

## 【0006】

図8に示すように、R A I D 装置は、複数の磁気ディスクユニット 91-1 ~ 91-4 と、これらを制御するディスクコントローラ 90 とからなる。図8では、4つの磁気ディスクユニット 91-1 ~ 91-4 からなるミラー構成の R A I



D装置を示している。

【0007】

即ち、磁気ディスクユニット91-1及び91-2により、論理ボリュームLM0が構成される。磁気ディスクユニット91-2には、磁気ディスクユニット91-1と同一の情報が記録される。磁気ディスクユニット91-3及び93-4により、論理ボリュームLM1が構成される。そして、磁気ディスクユニット91-4には、磁気ディスクユニット91-3と同一の情報が記録される。この構成は、1つの論理ディスクが2つの物理ディスクで構成されている例を示す。

【0008】

このため、磁気ディスクユニット91-1が故障しても、磁気ディスクユニット91-2を使用して、論理ボリュームLM0へのアクセスが可能である。磁気ディスクユニット91-3が故障しても、磁気ディスクユニット91-4を使用して、論理ボリュームLM1へのアクセスが可能である。

【0009】

磁気ディスクユニット91-1、91-2両方とも、正常に動作している場合には、どちらか一方の磁気ディスクユニットにアクセスする。この方法としては、従来は、片方の磁気ディスクユニットのみを使用することが行われていた。

【0010】

【発明が解決しようとする課題】

しかしながら、複数の上位装置に接続されている場合、上位から同一の論理ボリュームに2つ以上のアクセス要求が発生する場合がある。このような場合に、一方のみの物理ディスクユニットを使用する方法では、一の物理ディスクユニットが、2つのオペレーションを同時に実行できないため、アクセス要求は待たされる。

【0011】

同様に、一の物理ディスクユニットに、複数の論理ボリュームが設定されている場合には、上位から一の物理ディスクユニットの一の論理ボリュームにアクセス要求が発生することによって、上位から一の物理ディスクユニットの他の論理ボリュームにアクセスが発生することがある。このような場合、上位が1つであ

っても、アクセス要求は待たされる。

【0012】

又、磁気ディスクユニットの動作は、上位装置の動作に比し、比較的長時間かかるため、上位から頻繁に、同一の論理ボリュームへのアクセス要求が発生すると、待たされるアクセス要求の数が多くなる。このため、アクセス要求からオペレーション終了までの時間がかかり、システムのアクセス速度が低下するという問題があった。

【0013】

本発明の目的は、論理ディスクを構成する複数の物理ディスクユニットを使用して物理ディスクユニットへのアクセス要求の偏りを防止するためのRAID装置及びそのアクセス制御方法を提供することにある。

【0014】

本発明の他の目的は、物理ディスクユニット間の負荷の自動調整を行うためのRAID装置及びアクセス制御方法を提供することにある。

【0015】

本発明の更に他の目的は、物理ディスクユニットの負荷の偏りによる性能低下を防止するためのRAID装置及びそのアクセス制御方法を提供することにある。

【0016】

【課題を解決するための手段】

図1は、本発明の原理図である。

【0017】

図1に示すように、本発明のRAID装置は、論理ボリュームLM0、LM1を配置するための複数の物理ディスクユニット11-0～11-4と、指定された論理ボリュームが配置された物理ディスクユニットをアクセスして、指定された論理ボリュームをアクセスするディスクコントローラ10とを有する。

【0018】

このディスクコントローラ10は、各物理ディスクユニット毎に、各物理ディスクユニットに要求されているオペレーションの数を格納するメモリ22と、指

定された論理ボリュームが配置された前記複数の物理ディスクユニットを選択し、且つ選択された複数の物理ディスクユニットの内、オペレーションの数が最小である物理ディスクユニットをアクセスするための制御手段21とを有する。

【0019】

又、本発明のアクセス制御方法は、指定された論理ボリュームを構成する複数の物理ディスクユニットを選択するステップと、選択された複数の物理ディスクユニットの内、前記物理ディスクユニットに要求されているオペレーションの数が最小である物理ディスクユニットをアクセスするステップとを有する。

【0020】

本発明は、各物理ディスクユニットに要求されているオペレーションの数を、物理ディスクユニット毎に、記憶しておき、このオペレーションの数に応じて、アクセス要求を処理する物理ディスクユニットを決定するようにしている。

【0021】

各物理ディスクユニットに要求されるオペレーションの数は、各物理ディスクユニットの処理中及び処理待ちのオペレーションの数である。従って、物理ディスクユニットの実際の負荷数を常時測定している。そして、同一の論理ボリュームを格納する複数の物理ディスクユニットのオペレーションの数（負荷数）を判断し、オペレーションの数が最小の物理ディスクユニットを選択している。

【0022】

即ち、一の論理ボリュームが配置された複数の物理ディスクユニットの各々の負荷数が均等になるように制御している。このため、物理ディスクユニットへのアクセス要求の偏りを防止することができ、アクセス速度を向上する。

【0023】

この点で、一定時間毎に、交互に物理ディスクユニットを選択することが考えられる。しかしながら、上位からのアクセス要求は、均等に到来するものではないから、この方法では、同一の論理ボリュームを格納する複数の物理ディスクユニットの各々の負荷数が均等になるように制御することは、困難である。

【0024】

本発明では、物理ディスクユニットの実際の負荷数を常時測定しているので、

複数の物理ディスクユニットの各々の負荷数が均等になるように制御することができる。

【0025】

【発明の実施の形態】

図2は本発明の一実施の形態の構成図、図3は図2における論理ボリューム構成テーブルの説明図、図4は図2におけるDM管理テーブルの説明図である。

【0026】

図2において、ディスクコントローラ10は、磁気ディスクコントローラで構成されている。ディスクコントローラ10は、チャネルアダプター回路20と、リソースマネージャー回路21と、テーブルストレージ22と、メインストレージ23と、デバイスアダプター回路24-0～24-3とを有する。

【0027】

チャネルアダプター回路20は、コンピュータ等の上位装置とコマンド/データのやりとりを行うためのものである。リソースマネージャー回路21は、資源の管理を行うため制御を行う。リソースマネージャー回路21は、マイクロプロセッサで構成されている。

【0028】

テーブルストレージ22は、制御のための各種のテーブルを格納する。テーブルストレージ22は、図3にて後述する論理ボリューム構成テーブル22-1と、図4にて後述するDM管理テーブル22-2とを格納する。

【0029】

メインストレージ23は、リード/ライトデータ等を格納するためのものである。デバイスアダプター回路24-0～24-3は、リソースマネージャー回路21からの依頼に応じて、デバイス（磁気ディスクユニット）11-0～11-3をアクセス制御するものである。

【0030】

デバイスアダプター回路24-0には、磁気ディスクユニット11-0、11-2-1が接続される。デバイスアダプター回路24-1には、磁気ディスクユニット11-1、11-3-1が接続される。

【0031】

デバイスアダプター回路24-2は、デバイスアダプター回路24-0の予備用の回路であり、同じく、磁気ディスクユニット11-0、11-2が接続される。デバイスアダプター回路24-3は、デバイスアダプター回路24-1の予備用の回路であり、同じく、磁気ディスクユニット11-1、11-3が接続される。

【0032】

磁気ディスクユニット11-0～11-3は、周知の磁気ディスク記憶デバイスで構成されている。磁気ディスクユニット11-0及び磁気ディスクユニット11-1は、論理ボリュームLM0を構成する。磁気ディスクユニット11-2及び磁気ディスクユニット11-3は、論理ボリュームLM1を構成する。

【0033】

この構成は、RAID-1のミラー構成である。ミラー構成では、各論理ボリュームLM0、LM1は、物理ディスクを多重化して、配置される。本実施例にあっては、同一の論理ボリュームには、2つの物理ディスクが配置される。即ち、1つの論理ディスクが、2つの物理ディスクで構成されている。

【0034】

チャネルアダプター回路20は、上位装置からアクセス要求を受けると、リソースマネージャー回路21に、デバイスアクセス要求を発生する。リソースマネージャー回路21は、デバイスアクセス要求を受けると、デバイスアクセスパスを決定して、そして、オペレーションをデバイスアダプター回路に依頼する。

【0035】

デバイスアダプター回路は、磁気ディスクユニット毎に、オペレーションをキューイングする。そして、デバイスアダプター回路は、キューイングした順に、磁気ディスクユニットをアクセスする。

【0036】

デバイスアダプター回路は、磁気ディスクユニットのアクセスを開始する時に、データ転送要求をリソースマネージャー回路21に発する。リソースマネージャー回路21は、メインストレージ23の領域を割り当て、デバイスアダプター

回路に、データ転送を許可する。これにより、デバイスアダプター回路は、磁気ディスクユニットのデータをメインストレージ23に転送する。

【0037】

デバイスアダプター回路のデータ転送終了通知があると、リソースマネージャー回路21は、チャンネルアダプター回路20に、データアクセスを許可する。これにより、チャンネルアダプター回路20は、リード処理なら、メインストレージ23のデータを上位装置に転送する。チャンネルアダプター回路20は、ライト処理なら、メインストレージ23のデータに、上位装置からのライトデータを書き込む。

【0038】

チャンネルアダプター回路20が、データ転送を終了すると、リソースマネージャー回路21は、リード処理なら、メインストレージ23の領域を解放する。又、リソースマネージャー回路21は、ライト処理なら、メインストレージ23のデータを磁気ディスクユニットにライトバック処理した後、メインストレージ23を解放する。

【0039】

このライト処理は、メモリ21にデータをステージングして、ライト処理する。ミラー構成においては、ライトバックは、同一の論理ボリュームを格納する一対の磁気ディスクユニットに対して行われる。

【0040】

図3に示すように、テーブルストレージ22の論理ボリューム構成テーブル22-1は、各論理ボリューム「0」～「255」のステータス及び構成DM番号を格納する。ステータスは、各論理ボリュームの構成定義情報と、ミラーリング情報である。ミラーリング情報は、2ビットで構成され、ミラーを構成する一対の磁気ディスクユニットの各々が正常か異常かを示す。2つの構成DM番号1、2は、各々ミラーを構成する一対の磁気ディスクユニットの番号を示す。

【0041】

図4に示すように、テーブルストレージ22のDM管理テーブル22-2は、各磁気ディスクユニット「0」～「255」のステータス、負荷数及び接続DA

番号を格納する。ステータスは、その磁気ディスクユニットが正常か異常かを示すデグレード情報と、各磁気ディスクユニットが接続された一対のデバイスアダプター回路の各々が、正常か異常かを示す情報とを有する。

【0042】

負荷数は、その磁気ディスクユニットに要求されているオペレーションの数を示す。即ち、その磁気ディスクユニットの処理中及び処理待ちのオペレーション数を示す。2つの接続DA番号1、2は、磁気ディスクユニットが接続された一対のデバイスアダプター回路の番号を示す。

【0043】

図5は、本発明の一実施の形態のアイドル処理フロー図、図6は、図5における要求処理フロー図、図7は、図6におけるデバイスパス選択処理フロー図である。

【0044】

図5乃至図7は、図2のリソースマネージャー回路21が実行する処理である。図5のアイドル処理から説明する。

【0045】

(S1) リソースマネージャー回路(以下、プロセッサという)21は、チャンネルアダプター回路20及びデバイスアダプター回路24-0~24-3からの処理要求を受け付ける。

【0046】

(S2) プロセッサ21は、処理要求がないと、ステップS1に戻る。

【0047】

(S3) プロセッサ21は、処理要求があると、図6に示す要求処理を実行した後、ステップS1に戻る。

【0048】

次に、図6の要求処理を説明する。

【0049】

(S5) プロセッサ21は、処理要求が、チャンネルアダプター回路(CA)20からのデバイスアクセス要求かどうか調べる。プロセッサ21は、処理要

求がチャネルアダプター回路（CA）20からのデバイスアクセス要求であると判定した場合には、図7に示すデバイスパス選択処理を実行した後、終了する。

【0050】

尚、チャネルアダプター回路（CA）20は、上位装置からアクセス要求を受けると、プロセッサ21に対し、デバイスアクセス要求を出力する。

【0051】

（S6）プロセッサ21は、処理要求がチャネルアダプター回路（CA）20からのデバイスアクセス要求でないと判定した場合には、処理要求がデバイスアダプター回路（DA）24-0～24-3からのデータ転送要求かどうかを判定する。

【0052】

プロセッサ21は、デバイスアダプター回路（DA）24-0～24-3からのデータ転送要求であると判定した場合には、メインストレージ（MS）23の領域を割り当てた後、デバイスアダプター回路（DA）にデータ転送を許可する。そして、終了する。

【0053】

尚、デバイスアダプター回路（DA）は、プロセッサ21からのオペレーションの依頼をキューイングした後、その依頼を処理する時に、データ転送要求を発する。

【0054】

（S7）プロセッサ21は、デバイスアダプター回路（DA）24-0～24-3からのデータ転送要求でないと判定した場合には、処理要求が、デバイスアダプター回路（DA）24-0～24-3からのデータ転送終了通知であるかを判定する。

【0055】

プロセッサ21は、デバイスアダプター回路（DA）24-0～24-3からのデータ転送終了通知であると判定すると、終了通知からデータ転送の終了した磁気ディスクユニット（DM）を判断した後、DM管理テーブル22-2のその磁気ディスクユニットの負荷数から「1」減じる。



【0056】

そして、チャンネルアダプター回路（CA）20に、データアクセスを許可した後、終了する。これにより、チャンネルアダプター回路20は、リード処理なら、メインストレージ23のデータを上位装置に転送する。チャンネルアダプター回路20は、ライト処理なら、メインストレージ23のデータに、上位装置からのライトデータを書き込む。

【0057】

尚、デバイスアダプター回路（DA）は、オペレーションを実行して、磁気ディスクユニットからデータをメインストレージ（MS）23に転送終了後、データ転送終了通知を発して、次のオペレーションの処理に移る。

【0058】

（S8）プロセッサ21は、デバイスアダプター回路（DA）24-0～24-3からのデータ転送終了通知でないと判定すると、処理要求が、チャンネルアダプター回路（CA）からのデータ転送終了通知であるかを判定する。

【0059】

プロセッサ21は、処理要求が、チャンネルアダプター回路（CA）からのデータ転送終了通知であると判定すると、リード処理なら、メインストレージ23の領域を解放した後、終了する。

【0060】

又、リード処理でない、ライト処理なら、ライトバック処理を行う。即ち、ミラーを構成する一対の磁気ディスクユニットを選択した後、メインストレージ23のデータを一対の磁気ディスクユニットに転送する。そして、終了する。

【0061】

次に、図6のデバイスパス選択処理について、図7の処理フローにより説明する。

【0062】

プロセッサ21は、チャンネルアダプター回路20からのアクセス要求を受けた時に、デバイスパスを選択し、且つそのデバイスパスでのオペレーションをデバイスアダプター回路に依頼する。

【0063】

(S10) 先ず、プロセッサ21は、テーブルストレージ22の論理ボリューム構成テーブル22-1を参照する。プロセッサ21は、指定された論理ボリュームのステータス情報を調べる。ステータス情報のミラーリング情報からミラーを構成する一対の磁気ディスクユニットの各々が正常か異常か調べる。

【0064】

ミラーリング情報が異常を示している場合には、ステップS14に進む。一方、ミラーリング情報が正常を示している場合には、ステップS11に進む。

【0065】

(S11) ミラーリング情報が正常を示している場合には、ミラーを構成する一対の磁気ディスクユニットの各々が正常である。従って、プロセッサ21は、論理ボリューム構成テーブル22-1の当該論理ボリュームを格納する一対の磁気ディスクユニット番号(DM1、DM2)を求める。

【0066】

そして、プロセッサ21は、DM管理テーブル22-2の当該一対の磁気ディスクユニット番号の各々の負荷数A1、A2を読み出す。そして、プロセッサ21は、一対の磁気ディスクユニットの負荷数A1、A2を比較する。

【0067】

この比較により、一方の磁気ディスクユニットの負荷数A1が、他方の磁気ディスクユニットの負荷数A2以上である場合には、一方の磁気ディスクユニットの負荷が重いことを示す。従って、負荷の軽い他方の磁気ディスクユニットを選択するため、ステップS15に進む。

【0068】

逆に、一方の磁気ディスクユニットの負荷数A1が、他方の磁気ディスクユニットの負荷数A2より少ない場合には、一方の磁気ディスクユニットの負荷が軽いことを示す。従って、負荷の軽い一方の磁気ディスクユニットを選択するため、ステップS12に進む。

【0069】

(S12) プロセッサ21は、DM管理テーブル22-2の当該磁気ディス

クユニット（DM1）のステータス情報を参照する。このステータス情報は、当該磁気ディスクユニット（DM1）が接続されたデバイスアダプター回路の状態を示す。

【0070】

プロセッサ21は、このステータス情報により、当該磁気ディスクユニット（DM1）を接続した一对のデバイスアダプター回路のいずれも異常と判定すると、当該選択した磁気ディスクユニット（DM1）にアクセス出来ない。このため、プロセッサ21は、論理ボリューム構成テーブル22-1のステータス情報のミラーリング情報に、当該磁気ディスクユニット（DM1）の異常を書き込む。そして、ステップS14に進む。

【0071】

（S13）プロセッサ21は、このステータス情報により、当該磁気ディスクユニット（DM1）が接続された一对のデバイスアダプター回路の両方又はいずれかが正常と判定すると、当該選択した磁気ディスクユニット（DM1）にアクセス出来る。このため、プロセッサ21は、当該選択した磁気ディスクユニット（DM1）を接続しているデバイスアダプター回路に選択した磁気ディスクユニットのオペレーションを依頼する。

【0072】

そして、プロセッサ21は、DM管理テーブル22-2の当該磁気ディスクユニット番号の負荷数A1に、「1」を加算する。そして、終了する。

【0073】

（S14）プロセッサ21は、論理ボリューム構成テーブル22-1のミラーリング情報から異常のある磁気ディスクユニットは、どちらかを判定する。プロセッサ21は、他方の磁気ディスクユニット（DM2）がのみ異常であると判定すると、ステップS12に進む。これにより、一方の磁気ディスクユニット（DM1）の選択が行われる。

【0074】

一方、プロセッサ21は、一方の磁気ディスクユニット（DM1）が異常であると判定すると、ステップS15に進む。これにより、他方の磁気ディスクユ

ニット（DM2）の選択が行われる。

【0075】

更に、プロセッサ21は、両方の磁気ディスクユニット（DM1、DM2）に異常であると判定すると、エラーの終了を行う。

【0076】

（S15）プロセッサ21は、DM管理テーブル22-2の当該磁気ディスクユニット（DM2）のステータス情報を参照する。このステータス情報は、当該磁気ディスクユニット（DM2）が接続されたデバイスアダプター回路の状態を示す。

【0077】

プロセッサ21は、このステータス情報により、当該磁気ディスクユニット（DM2）を接続した一対のデバイスアダプター回路のいずれも異常と判定すると、当該選択した磁気ディスクユニット（DM2）にアクセス出来ない。このため、プロセッサ21は、論理ボリューム構成テーブル22-1のステータス情報のミラーリング情報に、当該磁気ディスクユニット（DM2）の異常を書き込む。そして、ステップS14に進む。

【0078】

（S16）プロセッサ21は、このステータス情報により、当該磁気ディスクユニット（DM2）が接続された一対のデバイスアダプター回路の両方又はいずれかが正常と判定すると、当該選択した磁気ディスクユニット（DM2）にアクセス出来る。このため、プロセッサ21は、当該選択した磁気ディスクユニット（DM2）を接続しているデバイスアダプター回路に選択した磁気ディスクユニットのオペレーションを依頼する。

【0079】

そして、プロセッサ21は、DM管理テーブル22-2の当該磁気ディスクユニット番号の負荷数A2に、「1」を加算する。そして、終了する。

【0080】

このようにして、DM管理テーブル22-2に、各磁気ディスクユニットの負荷数（オペレーション数）を格納している。そして、これを、オペレーションの

依頼毎に、加算し、且つ転送終了（アクセス終了）毎に、減算するため、常に、各磁気ディスクユニットの処理待ち及び処理中を示すオペレーション数（負荷数）を把握できる。

【0081】

このオペレーション数を、プロセッサ（リソースマネージャ回路）が、デバイスパスの選択の時に、参照して、ミラー構成である一对の磁気ディスクユニットの内、オペレーション数の小さい磁気ディスクユニットを選択する。このため、一对の磁気ディスクユニットの負荷数が均等となり、アクセス速度が向上する。

【0082】

これに対し、一定時間毎に、交互に磁気ディスクユニットを選択することが考えられる。しかしながら、上位からのアクセス要求は、均等に到来するものではなく、しかも1回のアクセスの時間も均一でない。このため、この方法では、同一の論理ボリュームを格納する複数の磁気ディスクユニットの各々の負荷数が均等になるように制御することは、困難である。

【0083】

同様に、アクセス要求毎に、交互に磁気ディスクユニットを選択することも考えられる。しかし、1回のアクセスの時間が均一でないため、一对の磁気ディスクユニットの各々の負荷数が均等になるように制御することはできない。

【0084】

これに対し、本発明では、各磁気ディスクユニットの負荷数を直接監視しているため、一对の磁気ディスクユニットの負荷数が均等になるように制御できる。

【0085】

又、従来は、磁気ディスクユニットの状態をデバイスアダプター回路が管理していた。そして、リソースマネージャ回路は、デバイスアダプター回路にオペレーションを依頼した後のデバイスアダプター回路からの応答により、選択した磁気ディスクユニットの異常を知るものであった。このため、従来では、選択した磁気ディスクユニットが異常の場合に、再度の選択処理及びオペレーションを必要としていた。

【0086】

この点で、この実施例では、論理ボリューム構成テーブルに、論理ボリューム毎に、一対の磁気ディスクユニットの状態を示すステータスを格納している。そして、リソースマネージャー回路が、論理ボリュームに対応する磁気ディスクユニットを選択する時に、このステータスを参照することにより、異常のある磁気ディスクユニットを知ることができる。

【0087】

このため、故障した磁気ディスクユニットの選択を未然に防止することができる。即ち、ミラー構成において、一の磁気ディスクユニットが故障であると、他の磁気ディスクユニットを、負荷数に係わらず、自動的に選択することができる。

【0088】

更に、この実施例では、DM管理テーブルに、その磁気ディスクユニットを接続したデバイスアダプター回路の状態を示すステータスを格納している。そして、リソースマネージャー回路が、論理ボリュームに対応する磁気ディスクユニットを選択する時に、このステータスを参照することにより、異常のあるデバイスアダプター回路を知ることができる。

【0089】

このため、故障したデバイスアダプター回路へのオペレーションの依頼を未然に防止することができる。尚、このDM管理テーブルのステータスは、リソースマネージャー回路からのオペレーションの依頼の時に、異常のあるデバイスアダプター回路は、エラーを通知してくる。これにより、DM管理テーブルの状態を更新することができる。

【0090】

上述の実施の態様の他に、本発明は、次のような変形が可能である。

【0091】

①前述の実施の態様では、論理ボリュームを2重化したミラー構成であるRAID-1で説明したが、論理ボリュームを3重化以上したものにも、適用できる。

【0092】

②物理ディスクユニットを、磁気ディスクユニットで説明したが、光ディスクユニット、光磁気ディスクユニット等を用いることもできる。

【0093】

以上、本発明を実施の形態により説明したが、本発明の主旨の範囲内で種々の変形が可能であり、これらを本発明の範囲から排除するものではない。

【0094】

【発明の効果】

以上説明したように、本発明によれば、次の効果を奏する。

【0095】

①各物理ディスクユニットのオペレーション数を監視し、且つオペレーション数の少ない物理ディスクユニットを選択するため、論理ボリュームを構成する複数の物理ディスクユニットの負荷数を均等にすることができる。

【0096】

②負荷の偏りを防止することができるため、アクセス速度を向上することができる。

【0097】

③各物理ディスクユニットのオペレーション数を監視しているため、物理ディスクユニットの動作速度に係わらず、負荷の偏りを正確に防止できる。

【図面の簡単な説明】

【図1】

本発明の原理図である。

【図2】

本発明の一実施の形態の構成図である。

【図3】

図2の実施の形態の論理ボリューム構成テーブルの説明図である。

【図4】

図2の実施の形態のDM管理テーブルの説明図である。

【図5】

図2の実施の形態のアイドル処理フロー図である。

【図6】

図5の実施の形態の要求処理フロー図である。

【図7】

図6の実施の形態のデバイスパス選択処理フロー図である。

【図8】

従来技術の説明図である。

【符号の説明】

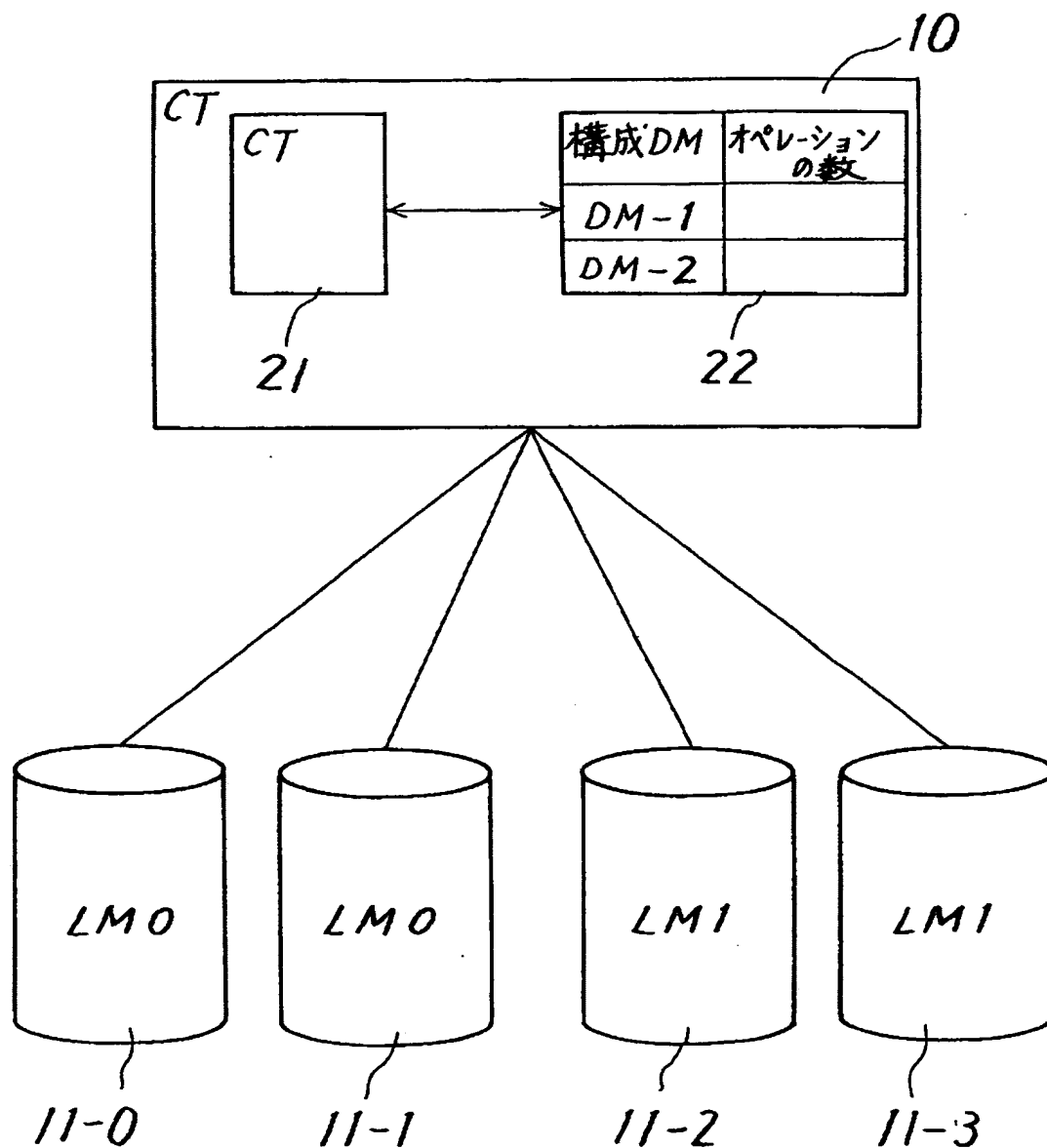
- 10、90 ディスクコントローラ
- 11-0～11-3 物理ディスクユニット
- 20 チャンネルアダプター回路
- 21 リソースマネージャ回路
- 22 テーブルストレージ
- 22-1 論理ボリューム構成テーブル
- 22-2 DM管理テーブル
- 23 データストレージ
- 24-0～24-3 デバイスアダプター回路
- LM0～LM1 論理ボリューム



【書類名】 図面

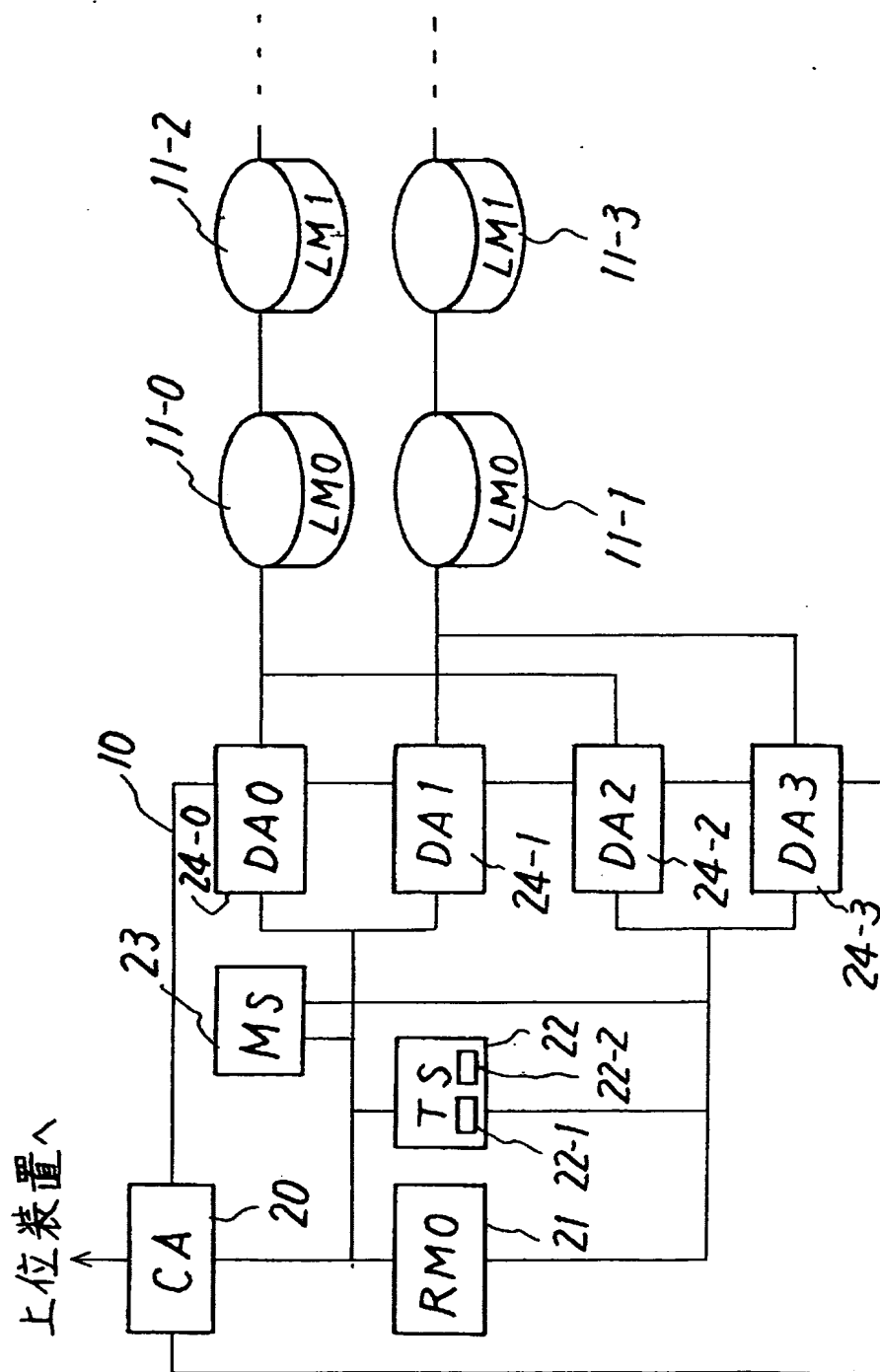
【図1】

# 原理図



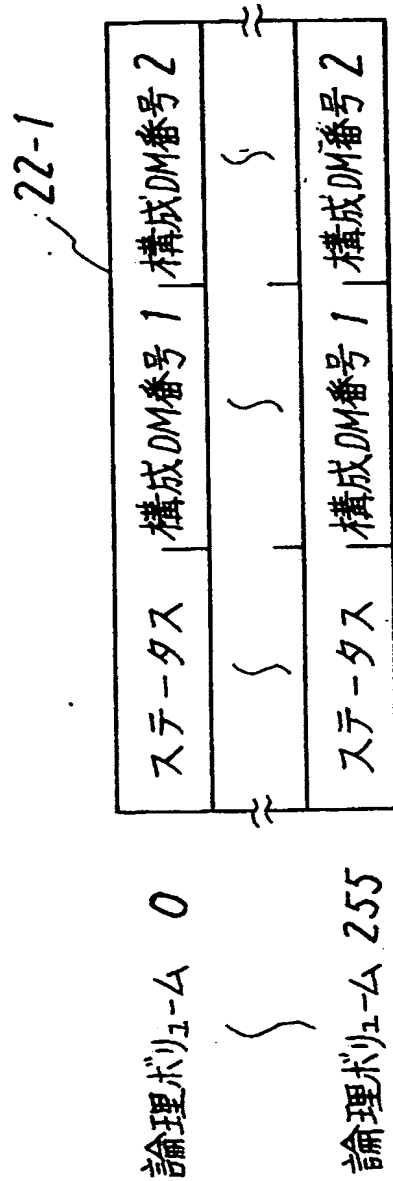
【図2】

構成図



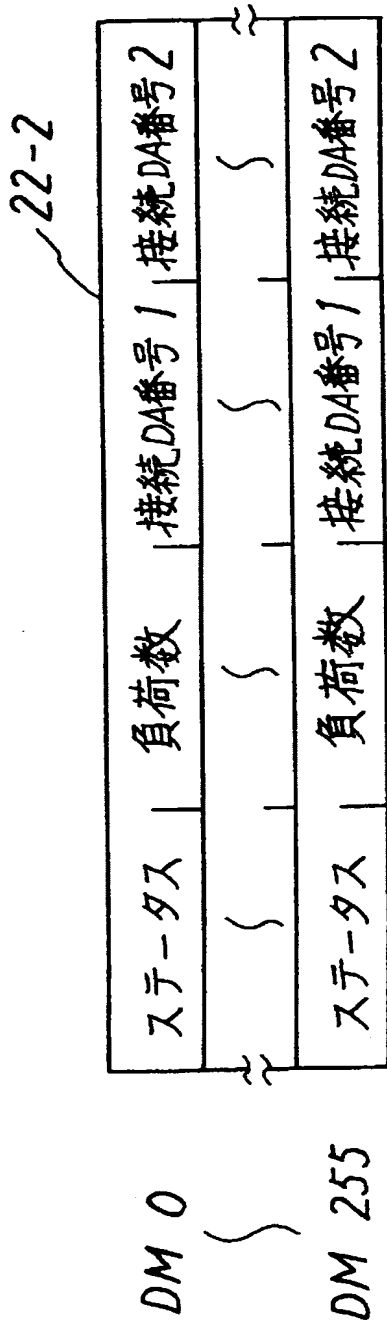
【図3】

論理ボリューム構成テーブルの説明図



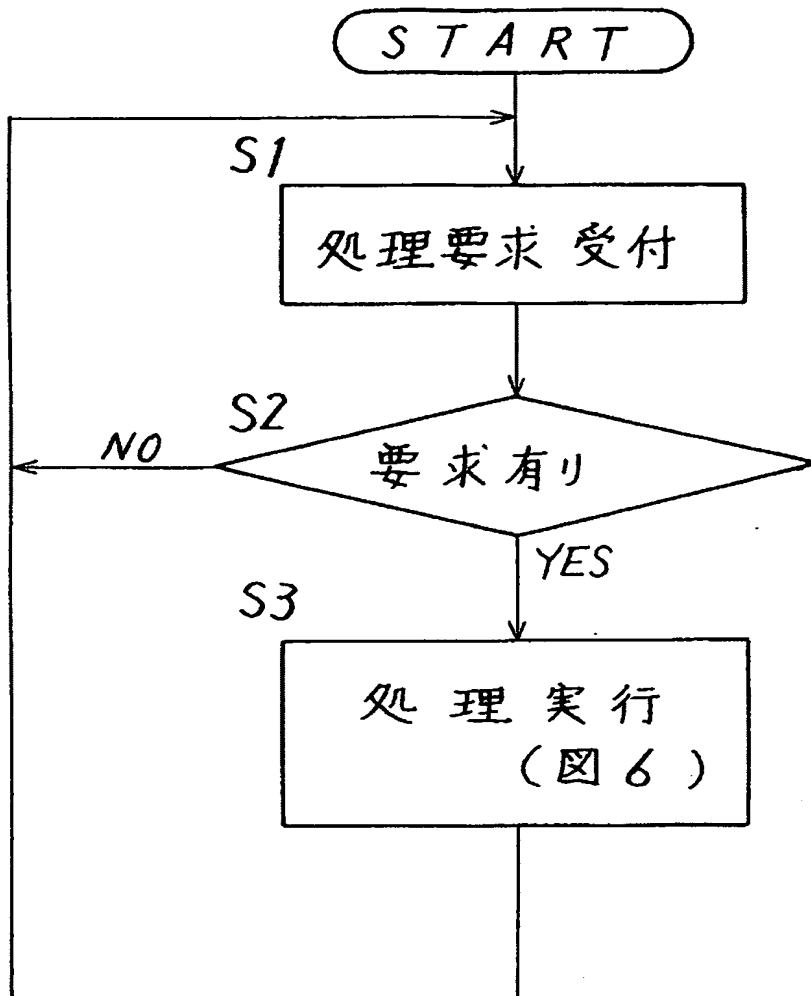
【図4】

DM管理テーブルの説明図



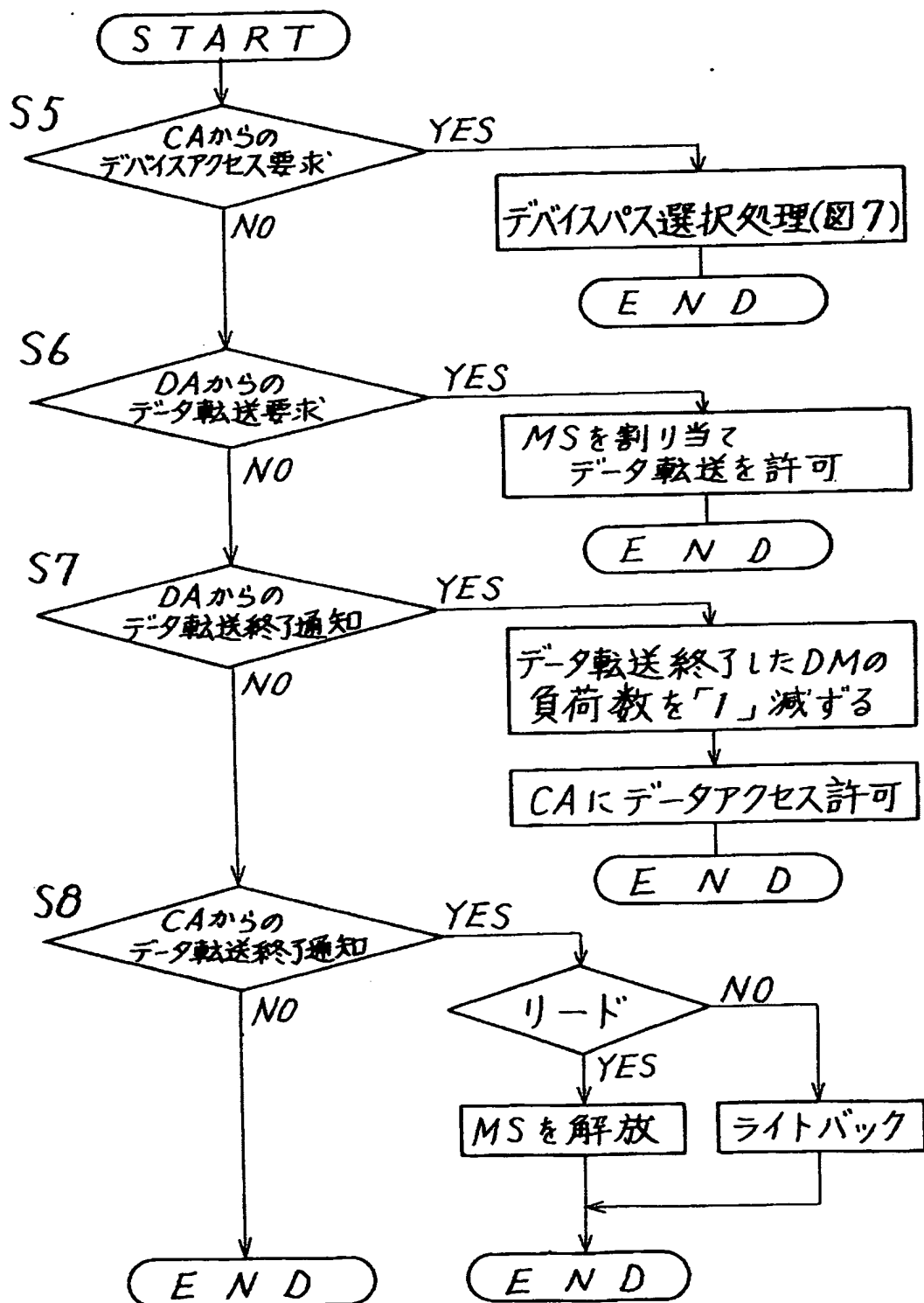
【図5】

# アイドル処理フロー図



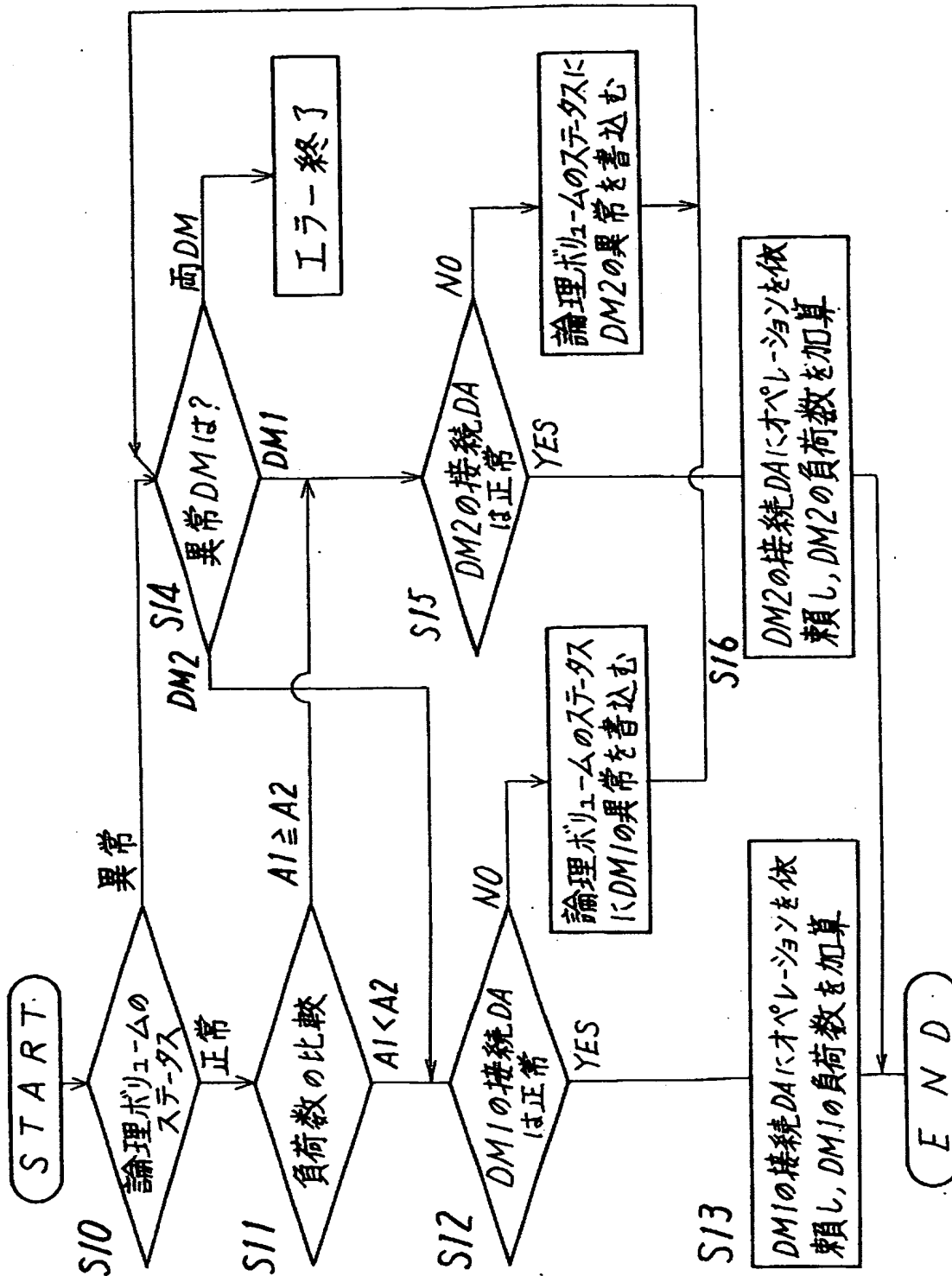
【図6】

# 要求処理フロー図



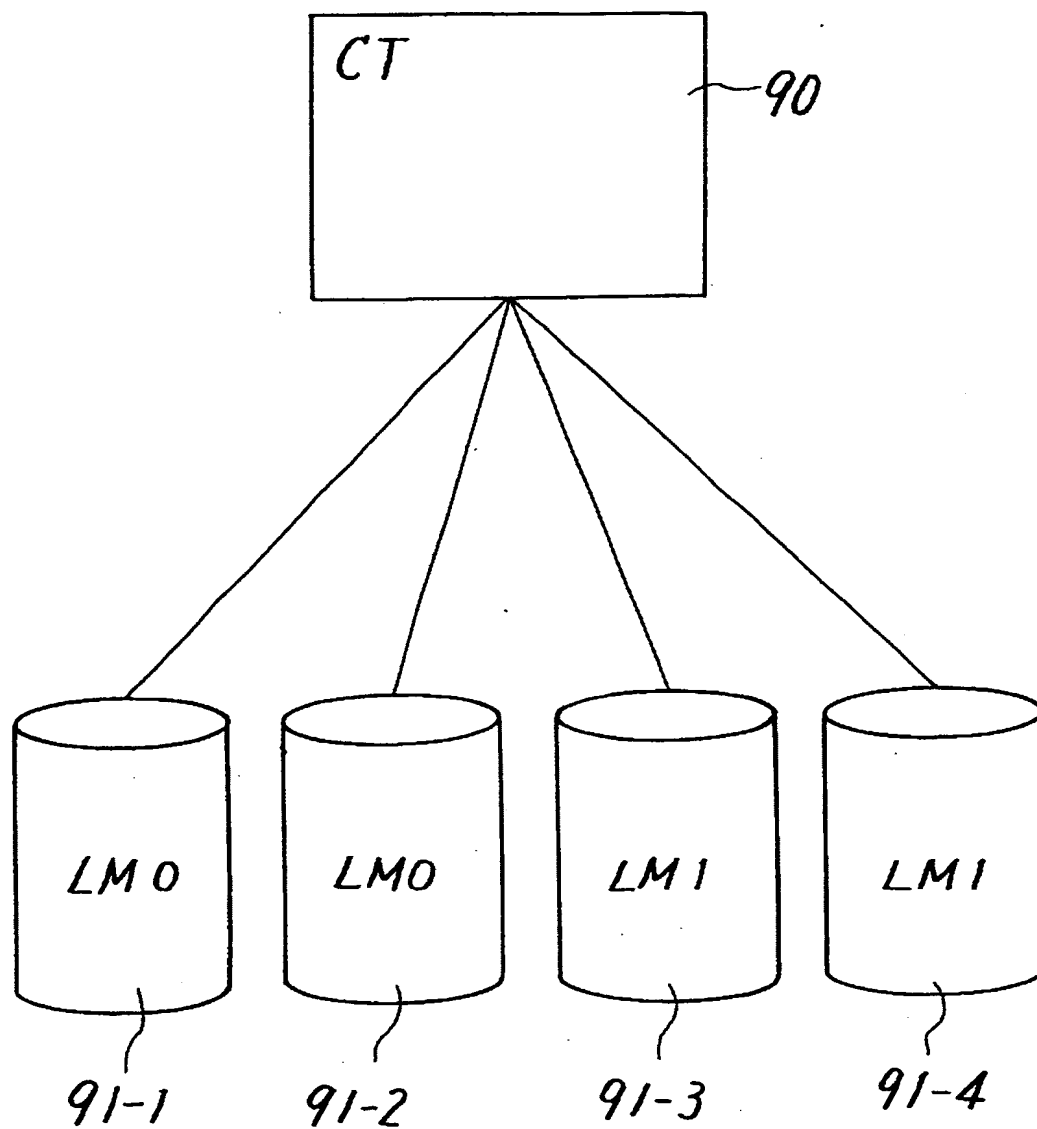
【図7】

デバイスパス選択処理フロー図



【図8】

従来技術の説明図





【書類名】 要約書

【要約】

【課題】 同一の論理ボリュームを複数の物理ディスクユニットで構成したRAID装置及びそのアクセス制御方法に関し、物理ディスクユニットのアクセスの偏りを防止する。

【解決手段】 一の論理ボリュームを配置するための複数の物理ディスクユニット11-0～11-3と、指定された論理ボリュームを構成する物理ディスクユニットをアクセスして、前記指定された論理ボリュームをアクセスするディスクコントローラ10とを有する。ディスクコントローラは、各物理ディスクユニットに要求されているオペレーションの数を格納するメモリ22と、指定された論理ボリュームが配置された前記複数の物理ディスクユニットを選択し、且つ選択された複数の物理ディスクユニットの内、前記オペレーションの数が最小である物理ディスクユニットをアクセスするための制御手段21とを有する。

【選択図】 図1

【書類名】 職権訂正データ  
【訂正書類】 特許願

<認定情報・付加情報>

【特許出願人】

【識別番号】 000005223  
【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号  
【氏名又は名称】 富士通株式会社

【代理人】

【識別番号】 100094514  
【住所又は居所】 神奈川県横浜市港北区新横浜3-9-5 第三東昇  
ビル3階 林・土井 国際特許事務所  
【氏名又は名称】 林 恒徳

【代理人】

申請人  
【識別番号】 100094525  
【住所又は居所】 神奈川県横浜市港北区新横浜3-9-5 第三東昇  
ビル3階 林・土井 国際特許事務所  
【氏名又は名称】 土井 健二

出 願 人 履 歴 情 報

識別番号 [000005223]

1. 変更年月日	1996年 3月26日
[変更理由]	住所変更
住 所	神奈川県川崎市中原区上小田中4丁目1番1号
氏 名	富士通株式会社